Conserv Genet (2017) 18:439–451
DOI 10.1007/s10592-016-0919-1

RESEARCH ARTICLE

Do dams also stop frogs? Assessing population connectivity of coastal tailed frogs (*Ascaphus truei*) in the North Cascades National Park Service Complex

Jared A. Grummer[1] · **Adam D. Leaché**[1]

Received: 6 July 2016 / Accepted: 23 December 2016 / Published online: 30 January 2017

Abstract We investigated the effects of three hydroelectric dams and their associated lakes on the population structure and connectivity of the coastal tailed frog, *Ascaphus truei*, in the North Cascades National Park Service Complex. Three dams were erected on the Skagit River in northern-central Washington state between 1924 and 1953 and subsequently changed the natural shape and movement of the Skagit River and its tributaries. We collected 183 frogs and tadpoles from 13 tributaries and generated a dataset of >2500 loci (unlinked SNPs) using double digestion restriction site-associated DNA sequencing (ddRADseq). An analysis of molecular variance (AMOVA) identified ~ 99% of the genetic variation within groups, and the remaining variation among groups separated by dams, or the Skagit River. All populations exhibited low F_{ST} values with a maximum of 0.03474. A "de novo" discriminant analysis of principal components revealed two populations with no geographic structure. However, testing groups that were partitioned *a priori* by the dams revealed distinctiveness of groups down-river of the lowest dam (Gorge Dam). Coalescent-based analyses of recent migration suggest that up to 17.3% of each population is composed of migrants from other populations, and an estimation of effective migration rates revealed high levels of migration heterogeneity and population connectivity throughout the study area. Our results suggest that although the populations down-river from Gorge Dam are distinguishable using SNPs, *A. truei* population connectivity is high throughout the North Cascades National Park Service Complex.

Keywords Migration · Dam · National Park · Ascaphus · Amphibian · SNP

> I went out in my alpine yard and there it was...hundreds of miles of pure snow-covered rocks and virgin lakes and high timber... Below, instead of the world, I saw a sea of marshmallow clouds.—Jack Kerouac on the North Cascades National Park, taken from *The Dharma Bums*

Introduction

Identifying patterns of genetic diversity across the landscape can help guide conservation efforts and mitigate negative impacts on wildlife populations due to human land-use activities (Dudley et al. 2005; Parmesan 2006). A variety of factors can cause populations to diverge genetically, including human-mediated landscape alterations (Pess et al. 2008; Sepulveda and Lowe 2009) and natural landscape features (Spear et al. 2005). In the Pacific Northwest (loosely defined as the area from northern California to British Columbia, and eastward into western Idaho), river dams and heterogeneous topography have led to population differentiation in both amphibian and fish species (Funk et al. 2005; Pess et al. 2008).

Freshwater resources across the globe are becoming increasingly stressed due to the rapidly growing human population. Across the U.S., more than 5 million kilometers

Electronic supplementary material The online version of this article (doi:10.1007/s10592-016-0919-1) contains supplementary material, which is available to authorized users.

✉ Jared A. Grummer
grummer@uw.edu

[1] Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Box 351800, Seattle, WA 98195-1800, USA

of streams and rivers are dammed, with only approximately 0.25% having any sort of protection (Benke 1990; Pringle et al. 2000). Unfortunately, the damming of riverine systems often has large negative consequences for the ecosystems in those areas (Richter et al. 2003). Some ecological consequences include population reduction and extirpation of migratory fish, range fragmentation, and increases in the number of exotic species (Pringle et al. 2000). Dams also affect terrestrial/riparian habitats and the species that occupy them by altering water levels and water availability in the areas both upstream and downstream of dams (Nilsson and Berggren 2000). Though poorly studied in this context, amphibians have the potential to provide insight into both the aquatic and terrestrial effects precipitated by damming riverine systems.

Habitat fragmentation, particularly as a result of human activities, is a major cause of the worldwide loss of alpha diversity (Fahrig 2003). It can also drive population genetic processes that can increase the probability of local extinction, such as increased inbreeding, loss of heterozygosity, and genetic drift (Curtis and Taylor 2003). In the Pacific Northwest, urban development and timber harvesting are two major factors that have contributed to habitat loss and fragmentation (Murphy and Hall 1981). The damming of rivers, which is a form of habitat fragmentation for aquatic organisms, has also greatly affected organismal populations in the Pacific Northwest. Washington state has nearly 1200 dams across its 39 counties, 98 of which are hydroelectric. The dams of the Elwha River on Washington's Olympic Peninsula, for instance, were in place for nearly a century until their removal in 2012, and these dams have had severe impacts on the migration and habitat of fish and other aquatic species (Duda et al. 2008; Pess et al. 2008).

The Skagit River Hydroelectric Project, which is owned by Seattle City Light, is a hydroelectric system spanning two counties (Skagit and Whatcom) in Washington State and is composed of three dams along the Skagit River. The Skagit River courses for about 200 km in the U.S., and the dams are located between 140 and 160 km upriver from its mouth. All three dams, and their associated lakes, are located within the North Cascades National Park Service Complex (NCNP). Out of the total ~89% of hydroelectric power that Seattle receives from all hydroelectric dams, the Skagit River dams provide ~20% (http://www.seattle.gov). The Gorge, Diablo, and Ross dams were erected in 1924, 1930, and 1953, respectively.

Given its abundance in the NCNP and potential for long-distance dispersal, we chose to focus on the coastal tailed frog, *Ascaphus truei*, for this study. This species has a distribution limited to the mesic areas of the Pacific Northwest. *Ascaphus truei* individuals can be locally abundant in first- and second-order streams, and are rarely found in stagnant water (e.g., ponds and lakes). Their restrictive physiology requirements for both lower temperatures and high moisture levels have been thought to lead to limited over-land dispersal (Claussen 1973; Brown 1975). In a mark-recapture study, Wahbe et al. (2004) captured a majority of *A. truei* individuals within 50m of streams, with a small number of individuals discovered between 50 and 100 m from the stream, indicating microhabitat preferences and limited dispersal from streams. However, this proximity to the stream was likely due to sampling methodology and different forest structure >50 m from sampled streams. Notwithstanding, Corn and Bury (1989) documented *A. truei* individuals as far as 1km from the closest stream. The results of the Corn and Bury (1989) study are in-line with more current research, which has uncovered high levels of gene flow at relatively large distances and population connectivity of *A. truei* populations at a scale of 25–30 km on Washington's Olympic Peninsula (Spear and Storfer 2008).

We assessed genetic connectivity and population structure in *Ascaphus truei* using single nucleotide polymorphism (SNP) data. A variety of methods have been developed in the past few years to acquire SNPs from non-model organisms (e.g., Elshire et al. 2011; Etter et al. 2011; Peterson et al. 2012), which are capable of producing hundreds or thousands of loci from across the genome. Furthermore, these methods require little knowledge of the genome, which eliminates the need to invest in genomic resource development. Large SNP datasets have the potential to provide precise estimates of population genetic parameters and increased statistical power for estimating population differentiation (Felsenstein 2006; Björklund and Bergek 2009) and therefore hold particular promise for identifying population structure caused by processes that have occurred in recent timescales.

In this study, we aimed to determine whether or not anthropogenic alterations to the landscape, hydroelectric dams and concomitant lakes in this case, affect population connectivity of *Ascaphus truei* in the NCNP. Because the Skagit River courses through a narrow canyon in our study site, dams have raised the water level some 123 m above the "natural" water level of the river in some areas, which not only separates creeks from opposite sides of the river that used to be much closer, but also places lentic water in between suitable *A. truei* habitat. This is the main reason why we expect to see population structure in this species caused by these dams: greater distances between suitable *Ascaphus* habitat due to habitat fragmentation. We collected nearly 200 individuals from 13 named creeks (and their associated tributaries) in areas around these dams and their associated lakes, to test whether populations of *A. truei* are structured by (a) the hydroelectric dams, (b) the Skagit River and the lakes created by the dams, or (c) a combination of the dams and the Skagit River. Our null hypothesis is that no population structure exists in this area.

Materials and methods

Sample collection

We opportunistically collected 196 *Ascaphus truei* individuals between 2012 and 2013 from 13 creeks and their tributaries in the NCNP, comprising 25 unique localities (Fig.1; under U.S. Department of the Interior and Washington Department of Fish and Wildlife permit nos. NOCA-2012-SCI-0044, NOCA-2013-SCI-0013, and RCW 77-32-240, WAC 220-20-045). Seven individuals were adults, 189 were larvae, and all that were retained as vouchers are deposited into the University of Washington's Burke Museum Herpetology and Genetic Resources Collections (see Supplemental Table 1 for UWBM voucher numbers and locality information).

DNA data collection

Genomic DNA (gDNA) was extracted from either liver, toe clip, or tail clip using the Qiagen DNeasy extraction kit (Qiagen, Valencia, CA). Total gDNA quality was assessed qualitatively through visualization on a 1% agarose gel and quantitatively with a Qubit fluorometer (Life Technologies, Carlsbad, CA). We generated sequence data using the double digestion restriction site-associated DNA sequencing (ddRADseq) technique developed by Peterson et al. (2012). Samples were first digested for eight hours at 37 °C with the restriction enzymes SbfI ("rare" 8bp restriction site sequence [5′ CCTGCAGG 3′]; New England Biolabs, Ipswich, MA) and MspI ("common" 4 bp restriction site sequence [5′ CCGG 3′]; New England Biolabs). The enzyme T4 DNA ligase (New England Biolabs) was then used to ligate barcoded oligonucleotides to each genomic DNA fragment (each barcode 5bp in length) that were unique to each row of individuals on the sequencing plate (eight unique barcodes total); each barcode appears at the 5′ end of each sequenced fragment (see Peterson et al. 2012 for more details). Individuals were then pooled, followed by a size selection step with the Blue Pippin (Sage Science, Beverly, MA) where all loci between 415 and 515 bp were retained. A final PCR step using Phusion Taq polymerase

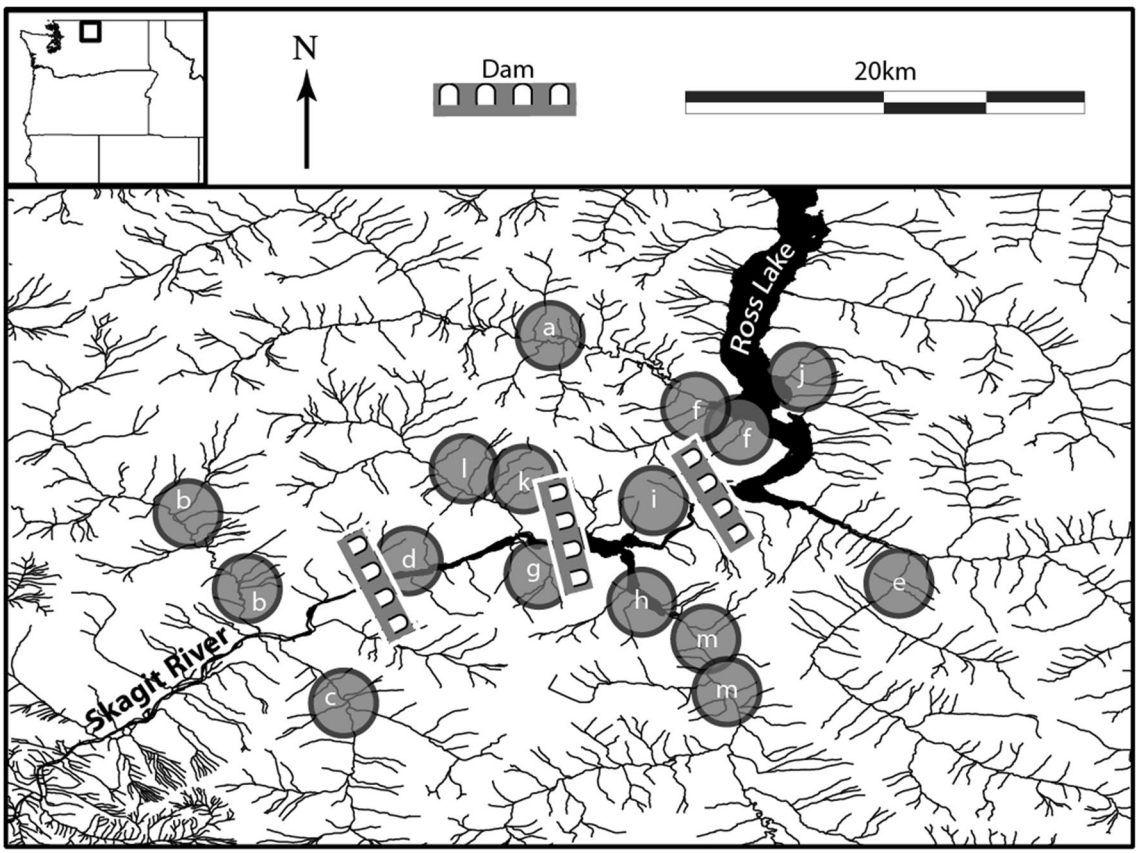

Fig. 1 Study area showing *Ascaphus truei* sample sites and waterways in the North Cascades National Park Service Complex. Letters correspond to the following creeks, with sample sizes following creek name: *a*—Big Beaver (16), *b*—Goodell (15), *c*—Newhalem (15), *d*—North Gorge (15), *e*—Panther (13), *f*—Pierce (14), *g*—Pyramid (20), *h*—Rhodes (15), *i*—Riprap(13), *j*—Roland (13), *k*—Sourdough (7), *l*—Stetattle (14), *m*—Thunder (13). Refer to Supplementary Table 1 for further sampling information

(New England Biolabs) with the following thermocycler conditions was conducted to amplify all loci and attach a 6bp index unique to each pool (12 unique indexes per plate) for sequence de-multiplexing: 98° for 0:30, (98° for 0:10, 58° for 0:30, 72° for 0:30) × 12 cycles, and a final 10:00 extension at 72 °C; each adapter sequence is attached to the 3′ end of each sequenced fragment. Ninety-six or 144 individuals (sometimes multiplexed with other organisms) were multiplexed across four separate sequencing runs at the University of California Berkeley QB3 Vincent J. Coates Genomics Sequencing Laboratory on an Illumina HiSeq 2500 with 50 bp single-end sequencing. Three creeks (Roland, Newhalem, North Gorge) were represented by sequence data from the same sequencing lane (e.g., all individuals from a creek were sequenced on the same lane), whereas sequencing effort was divided across 2–4 lanes for all other creeks (e.g., individuals from the same creek sequenced on different lanes).

DNA data assembly

Raw Illumina reads were processed with the program pyRAD v3.0.5 (Eaton 2014) to generate alignments of phased SNPs. We chose to use pyRAD because it can accommodate insertions/deletions in loci that programs such as STACKS (Catchen et al. 2011) cannot. Samples were first de-multiplexed based on unique barcode-index combinations, then sequence "clusters" were generated by pyRAD using the programs VSEARCH (Rognes et al. 2016) and MUSCLE (Edgar 2004). Sequence reads were discarded after de-multiplexing if they had ≥4 bp with a Phred quality score <20. After de-multiplexing and removing the restriction overhang, the reads were 39 bp in length. Reads were first clustered within individuals into loci that were ≥90% similar, and then across individuals using the same threshold. Loci were retained for a given individual if they had a minimum sequencing depth of 10× (loci often had a mean sequencing depth > 20×). PyRAD uses a paralog filter, which represents the maximum percentage of individuals allowed to have a heterozygous base (IUPAC "ambiguities") at a given site. A higher value for the paralog filter allows more heterozygotes at any given position that could be due to (a) fixed allelic differences, or (b) sequence polymorphism, both of which can appear the same due to sequence reads containing both alleles. For our final datasets, we set this value to 90%, meaning ≤90% of the individuals at a given locus (putatively unlinked SNP) could share a heterozygous SNP, because we expect that heterozygosity can occur at a high frequency at this limited spatial scale. We note that using a ≤25% heterozygosity cutoff did not drastically change dataset size (~10% reduction), and exploratory analyses revealed similar results to those presented with the ≤90% heterozygosity dataset (Supplemental Fig. 1). We compiled two datasets that differed with respect to the amount of missing allowed: one contained no missing data (e.g., 100% complete), and the second dataset allowed up to 50% missing data for any locus.

Identifying genetic subdivision

We tested three *a priori* hypotheses (along with the null hypothesis of no genetic structure) to determine which geographical feature (including dams), if any, is responsible for causing genetic subdivision between *A. truei* populations: genetic subdivision caused by (a) three hydroelectric dams, (b) the Skagit River, or (c) a combination of the dams and the Skagit River (Table 1).

We first tested for parentage and relationships among samples using the program Colony (Wang and Santure 2009; Jones and Wang 2010). Colony implements a maximum-likelihood method to inferring sib-ship, and has been shown to outperform other parentage inference methods (Harrison et al. 2013). We considered individuals to be full sibs that were inferred with a probability >95% (Hauser et al. 2011). We also tested for a correlation between genetic and geographic distances (isolation by distance) with a Mantel test in the program Adegenet (Jombart 2008; Jombart et al. 2010; Jombart and Ahmed 2011). The significance of isolation by distance was tested by creating a null distribution (an absence of spatial structure; 1000 replicates) and comparing the empirical value to this distribution. We assessed genetic variation among our pre-defined groups using an analysis of molecular variance (AMOVA) in Arlequin (v3.5; Excoffier and Lischer 2010). A locus-by-locus AMOVA was performed on the 100% complete dataset (results from all loci were combined for the final result). We also used Arlequin to calculate population-pairwise F_{ST} values, which were done with 1000 permutations to test for statistical significance of population differentiation.

We estimated the number of populations (k) using a discriminant analysis of principal components (DAPC) on the 100% complete dataset in the R package Adegenet (Jombart 2008; Jombart et al. 2010; Jombart and Ahmed 2011). The program first transforms the SNP dataset through a principal component analysis (PCA), then a discriminant analysis (DA) is performed on the output of the PCA analysis. Generally, the DAPC method seeks to maximize between-group genetic variation while minimizing within-group variation, and has the potential benefit over other population clustering methods of making no assumptions about the underlying population genetic model. We first performed a "de novo" analysis without individuals assigned to populations, and we chose the optimal k value based on the Bayesian Information Criterion (BIC) of the likelihood

Table 1 Population structure hypotheses tested in our analyses

Hypothesis	# Groups	Big Beaver	Panther	Pierce	Roland	Rhodes	Riprap	Thunder	North Gorge	Pyramid	Sourdough	Stetattle	Goodell	Newhalem
By dam	4	1	1	1	1	2	2	2	3	3	3	3	4	4
By river	2	1	2	1	2	2	1	2	1	2	1	1	1	2
By dams and river	8	7	8	7	8	6	5	6	3	4	3	3	1	2

Total number of groups tested under each hypothesis is shown, where numbers indicate group assignment that correspond to Fig. 2

score associated with each k iteration ("find.clusters" command). We then assigned individuals to populations based on the groups defined in Table 1 and explored population structuring under these three hypotheses. For all analyses, we used 60 principal components, which is approximately one-third the number of individuals (Jombart 2008).

We also used the program Admixture (Alexander et al. 2009) to determine k. This program is similar to the more popular program Structure (Pritchard et al. 2000) in that both approaches model the probability of the observed genotypes using ancestry proportions and population allele frequencies. However, one difference is that unlike Structure, which utilizes a Bayesian algorithm, Admixture uses a maximum likelihood approach that differs in how the optimal k value is selected. Specifically, the Evanno et al. (2005) method widely used in Structure analyses cannot evaluate $k = 1$. In contrast, the cross-validation method employed in Admixture can evaluate $k = 1$. We ran Admixture (Alexander et al. 2009) on our 50% complete dataset, running ten replicate analyses with unique starting seeds to ensure consistency of results.

Estimating recent migration rates

The human-mediated habitat change that we assessed in this study was very recent, within the past ~70 years. This amount of time equates to approximately 8–40 *A. truei* generations, given an estimated generation time of 2–8 years for this species (Bury and Adams 1999; Nielson et al. 2001). To estimate recent migration rates over the last several generations, we used the Bayesian program BayesAss v3.0.4 (Wilson and Rannala 2003). This program estimates the proportion of each population that are immigrants from each of the other populations in the system (e.g., asymmetric migration rates), in addition to estimating the total number of non-migrants, and first- and second-generation migrants. A benefit of this program over others that estimate migration is that it relaxes many population genetic assumptions such that the populations do not have to be at equilibrium. Importantly, genotype frequencies can deviate from Hardy-Weinberg equilibrium within populations. We ran four replicate analyses on our 50% complete dataset, which was partitioned in four different ways: three analyses where individuals were assigned to groups based on Table 1, and the fourth where individuals were assigned to the creek in which they were sampled. Each was run for 10^8 generations, and the first 2×10^7 generations were discarded as burn-in (with "mixing" parameters −a 0.4 −f 0.1 −m 0.2). Convergence was visually assessed in Tracer v1.5 (Rambaut and Drummond 2007).

Secondly, we estimated migration rates using the "estimating effective migration surfaces" (EEMS) method (Petkova et al. 2016). This method is based upon a stepping

stone model in which migration is allowed between neighboring demes in a grid, the density of which the user specifies. This approach assesses genetic connectivity across the landscape in a way that makes it conceptually related to methods that utilize circuit theory (Hanks and Hooten 2013). Migration rates are adjusted such that the genetic differences expected under the model are close to the genetic differences observed in the data. These estimates are then interpolated across the landscape to produce the estimated effected migration surface. This method requires n sites >>n individuals, so we used the larger (50% complete) data matrix that had 2341 loci after removing non-biallelic sites. This method produces *effective* migration rates that show genetic similarities across a landscape, not the typical N_em of some other methods. If genetic similarities decay faster in the observed data than expected, a lower effective migration rate will be inferred for that area (e.g., a barrier). We experimented with the number of demes (grid density) and ultimately used 100 demes because this provided a sufficient grid density for inferring migration rates across the study area and produced repeatable results across replicate runs. The analysis was run for 5×10^7 generations with 10^7 generations as burn-in and the chain state was saved every 50,000 generations. Convergence was assessed by inspecting concordance across replicate runs with different starting seeds, in addition to examining trace plots of the MCMC for stationarity.

Results

Population and genetic structure

After removing 13 samples (seven due to poor genomic DNA quality and six due to poor DNA sequences), our final datasets consisted of 183 individuals and 2537 or 211 loci (SNPs) for the 50 and 100% complete matrices, respectively; the number of individuals per creek ranged from 7–20 (Fig. 1; Table S1). Parentage analyses identified 5 full-sib pairs with > 95% confidence, where each sibling-pair was from the same creek. Removal of one individual from each full-sib pair for analyses (five individuals in total) did not change analysis results, so we present results that include all individuals. We did not detect any significant signal of isolation by distance in our dataset ($p = 0.10$; Supplemental Fig. 2). Our AMOVA results revealed that the majority of genetic variation is found within groups (>90%), e.g., not between pre-defined groups (Table 2). Whether the data were divided into groups separated by the dams, the Skagit River, or the dams and Skagit River made no difference in among-group genetic variation, which was <1% in all cases (Table 2).

Table 2 Results from AMOVA analyses on the 100% complete dataset

Source of variation	Percentage of variation		
	By dams	By river	By dams and river
Among groups	0.82	0.16	1.00
Within groups	99.18	99.84	99.00

Refer to Table 1 for assignments of individuals/creeks to groups

Estimates of F_{ST} between groups ranged from 0.00 to 0.03474 (Tables 3, 4, 5), indicating little differentiation between *a priori* defined groups. However, in spite of these low values, the majority of pairwise population differentiation tests were significant at $p < 0.05$ (Tables 3, 4, 5).

The results from our DAPC analyses are shown in Fig. 2. For the complete dataset, the optimal k value for the "*de novo*" analysis was $k = 2$. (Supplemental Fig. 3; see Supplemental Fig. 4 for $k = 3$ results). It is interesting to note that when using the 50% complete dataset (2537 loci), $k = 4$ was the best grouping based on BIC score (though only slightly better than $k = 3$; Supplemental Fig. 5, 6). There appears to be no clear geographic structure for any of the populations identified in the de novo analyses (Fig. 2a), nor any correlation with experimental conditions (i.e., sequencing plate or level of missing data, etc.).

When partitioned by dam, populations below Gorge Dam appear distinct (Fig. 2b). The populations above Ross Dam are also distinct, although with a fair amount of genetic overlap with central populations (Fig. 2b). Admixture is greatest in the central populations (Ross to Diablo and Diablo to Gorge dam stretches; Fig. 2b). Partitioning individuals by the Skagit River provided moderate genetic differentiation between these two clusters (Fig. 2c). And lastly, partitioning individuals by regions isolated by both dams and the Skagit River resulted in clear differentiation of the populations west of the Gorge Dam and both north and south of the Skagit River (Fig. 2d). Individuals from the Big Beaver drainage north of Ross Dam and west of the Skagit River also showed some distinctiveness from the other groups.

We conducted randomization tests to determine the likelihood of detecting population structure by chance. Though a characteristic of the data and its informativeness, this is largely a function of the congruence of the actual population structuring (as seen in the genetic data) with pre-defined groups. We randomized the assignments of individuals to two, four, and eight groups (to mimic the number of groups in our a priori hypotheses) and repeated the DAPC analyses. We used the same overall number of individuals assigned to each group as in our empirical data,

Table 3 Pairwise F_{ST} results when individuals were partitioned by dams

	Dam section			
	North of Ross	Ross to Diablo	Diablo to Gorge	South of Gorge
North of Ross	0			
Ross to Diablo	**0.00914**	0		
Diablo to Gorge	**0.00768**	0.00217	0	
South of Gorge	**0.01882**	**0.00893**	**0.01207**	0

Bold values indicate significance at $p < 0.05$; all other values are insignificant

Table 4 Pairwise F_{ST} results when individuals were partitioned by the Skagit River

	Side of river	
	North/West of Skagit	South/East of Skagit
North/West of Skagit	0	
South/East of Skagit	**0.00192**	0

The single bold value indicates significance at $p < 0.05$

and conducted ten independent replicates for each assignment scheme. For each replicate, we calculated the average proportion of individuals that were assigned to their correct group. If our hypotheses were supported, we would expect to see a high percentage of each individual's genotype (and therefore, a high average percentage across all individuals) assigned to its pre-defined population.

When pre-assigned to two groups, 68 and 66% of the individuals were correctly assigned to groups 1 and 2 across all ten replicates, respectively (Table 6; Supplemental Fig. 7). As opposed to the randomized groupings, the empirical groupings do a better job at discriminating the individuals, with 72 and 73% of genotypes correctly assigned to their pre-defined groups (Table 6). When assigned to four groups, the empirical data do better at assigning individuals to 3/4 groups (though only 1% worse in the 4th group), particularly with the group composed of individuals from Goodell and Newhalem Creeks (Group 4; Table 6). A similar pattern is seen with the $k = 8$ groupings, where the empirical assignments do better than the randomized assignments for the majority of groups, with the Goodell and Newhalem individuals (Groups 1 and 2) assigned to their respective groups with high percentages.

Our Admixture (Alexander et al. 2009) analysis identified the most likely number of populations as $k = 1$ (Supplemental Fig. 8). These results were stable across all 10 replicate runs.

Table 5 Pairwise F_{ST} results when individuals were partitioned by dams and the Skagit River

	Dam and river partition							
	North of Ross, West of Skagit	North of Ross, East of Skagit	Ross-Diablo, West Skagit	Ross-Diablo, East of Skagit	Diablo-Gorge, North of Skagit	Diablo-Gorge, South of Skagit	West of Gorge, North of Skagit	West of Gorge, South of Skagit
N Ross, W Skagit	0							
N Ross, E Skagit	**0.00857**	0						
Ross-Diablo, W Skagit	**0.01904**	0.00	0					
Ross-Diablo, E Skagit	**0.01488**	**0.00755**	0.00603	0				
Diablo-Gorge, N Skagit	**0.01877**	**0.00806**	0.00261	**0.00674**	0			
Diablo-Gorge, S Skagit	**0.00757**	0.00529	0.00195	0.00	0.00036	0		
W Gorge, N Skagit	**0.02633**	**0.03474**	**0.03057**	**0.02326**	**0.02048**	**0.01255**	0	
W Gorge, S Skagit	**0.02203**	**0.01906**	0.01247	0.00878	**0.01334**	0.00061	**0.0159**	0

Bold values indicate significance at $p < 0.05$; all other values are insignificant

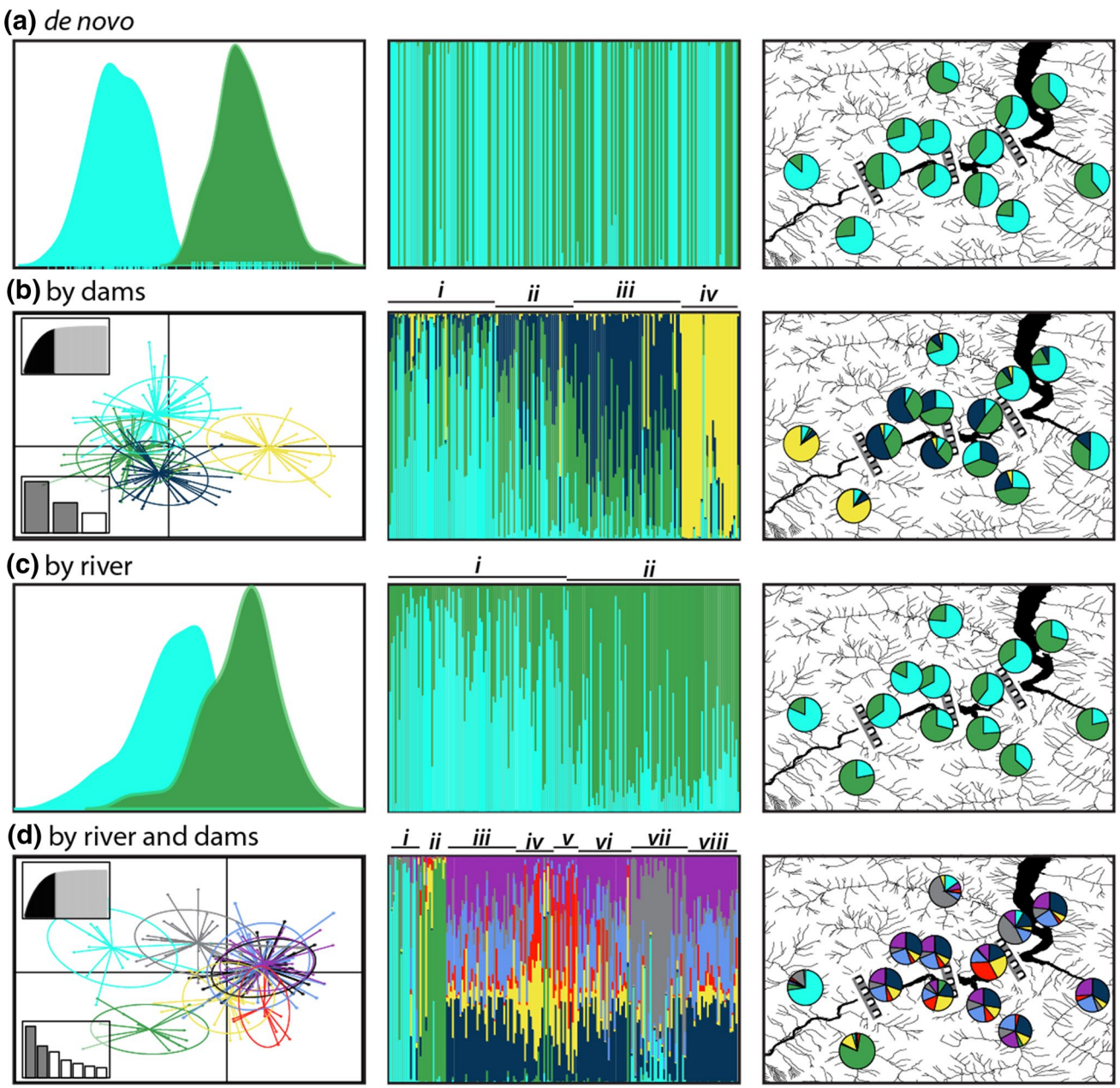

Fig. 2 Discriminant analysis of principal components (DAPC) results for three different hypotheses of genetic subdivision (**b**–**d**) along with de novo clustering (**a**). The left-hand column shows the DAPC plots, middle column shows the population assignments of individuals to each respective cluster, and the right column shows the percentage of individuals from each creek assigned to each cluster. *Insets* in the left column of rows **b** and **d** indicate number axes retained in principal component (*top-left*) and discriminant analyses (*bottom-left*). Roman numerals above figures in the middle column correspond with population assignments in Table 1. (Color figure online)

Migration rates

For our BayesAss results, variation in the posterior mean migration estimates across four replicates of each geographic partitioning scheme (i.e., hypotheses in Table 1) was low (often <0.000*x*), in spite of low effective sample size (ESS) values for the overall log-probability of each analysis (results not shown). Standard deviation of the posterior mean migration estimates were also low (<0.06); the migration estimates presented here are from a single analysis for each hypothesis tested. Migration rates inferred from the 100% complete dataset were similar to those inferred with the 50% complete dataset, with migration rates being slightly higher in the 100% complete analyses (results not shown). Migration rates across all geographic partitioning schemes were relatively high (>0.0075,

Table 6 Average assignment probabilities for *Ascaphus truei* using population clusters that were assigned randomly versus using the empirical hypotheses

Group	$k = 2$		$k = 4$		$k = 8$	
	Random	Empirical	Random	Empirical	Random	Empirical
1	68	72	54	66	28	73
2	66	73	45	44	29	79
3	–	–	50	55	39	31
4	–	–	39	83	32	27
5	–	–	–	–	18	33
6	–	–	–	–	34	21
7	–	–	–	–	33	53
8	–	–	–	–	24	24

Refer to Table 1 for individual assignments to group number

Supplemental Tables 2–5; note that values are expressed as proportion of the population, not N_em). The highest migration rate observed was from the population north and west of the Skagit River (Big Beaver Creek) into the population on the opposite side of the river (Roland Creek; ~17%; Supplemental Table 4). The lowest migration rates were those between populations partitioned by both dam and river (lowest value of 0.76%; Supplemental Table 5).

The EEMS analysis indicated a high level of migration heterogeneity throughout the study area (Fig. 3). A high level of migration was inferred between populations immediately above and below Ross Dam (Pierce and Riprap, respectively; Fig. 3). Similarly, population connectivity is high in the central portion of the study area across the Skagit River and between the Pyramid, Rhodes, and Sourdough populations. On the other hand, EEMS identified barriers to gene flow in two cases that correspond to large mountains. One barrier is in the north between Big Beaver and Sourdough/Stetattle populations, and the second is in the southeastern portion of the study site between Panther and Thunder Creeks. Above Ross Dam, Ross Lake appears to be a barrier between individuals from Pierce and Roland Creeks. The last migration barrier inferred by EEMS coincides with the Gorge Dam, which appears to decrease up- or down-river movement.

Discussion

Even though the Skagit River Hydroelectric Project was completed between the years of 1921 and 1953, the North Cascades National Park was not established until 1968. And in spite of the fact that *A. truei* is listed as a "species of concern" by the Washington Department of Natural Resources

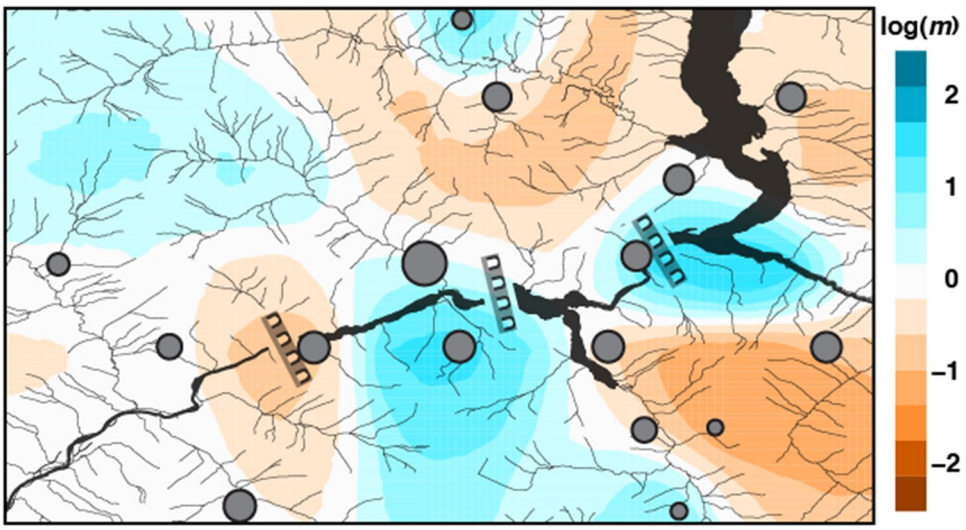

Fig. 3 EEMS results showing estimated effective migration rates for *Ascaphus truei* in the North Cascades National Park Service Complex. Sampling localities are shown as *grey circles* and are proportional in size to the number of individuals. *Darker blues* indicate high estimated migration rates, whereas *dark orange* indicates low migration rates, relative to the overall migration rate across the entire area. Thus, in the log scale on the right, a value of 2 corresponds to effective migration rates that are 100× higher than the average rate. These values do not indicate N_em. (Color figure online)

(due to logging and other forms of habitat destruction; http://www.dnr.wa.gov), it is listed as a species of "least concern" by the IUCN (http://www.iucnredlist.org/). There is relatively little to no support for genetic structuring among the majority of *A. truei* populations sampled in our study. However, we found evidence for weak population structuring in the North Cascades National Park Service Complex associated with Gorge Dam (the furthest down-river and oldest dam) and the Skagit River/lakes (Figs. 2, 3). This result indicates that geographic structure should be considered in future tailed frog conservation/management decisions in this area.

Ability to detect recent cessation in gene flow

The SNP dataset presented here was able to detect population structure associated with the oldest dam (Gorge Dam), but unable to detect similar structure across the more recently established dams. In 1921, construction of the Gorge Dam was initiated; this dam is also the furthest down-river of the three. The next dam up-river, Diablo, was completed in 1930. And finally, Ross Dam, which is the furthest up-river and largest of the three (166 m tall), was completed in three stages between 1940 and 1953. Overall, there has been little time for genome-wide mutations to accumulate due to genetic drift as a result of the dams. We estimate that only ~8–40 *Ascaphus* generations have elapsed since the dams were built 60–80 years ago. The generation time for *A. truei* is estimated to be ~2–8 years (Bury and Adams 1999; Nielson et al. 2001) and this long generation time is correlated with slow evolutionary and metabolic rates, at least for mitochondrial DNA (Martin and Palumbi 1993).

The DAPC did well at assigning individuals to their pre-assigned empirical groups better than randomized assignments in a majority of the groupings (hypotheses) we explored (Table 6). However, analyses on the randomized assignments also "correctly" assigned individuals to their pre-defined groups more often than would be expected. This is likely due to weak genetic structure across most of the landscape (except the distinctiveness of the Goodell and Newhalem Creek individuals), and the methodology of the DAPC that seeks to maximize differences between groups.

Both population assignment (DAPC) and effective migration rate (EEMS) analyses indicated some level of population structuring and areas of reduced gene flow in this system. It is interesting to note that EEMS detected strong barriers to migration that coincide with two large mountains: (1) between samples from Rhodes/Thunder Creeks and Panther Creek, and (2) between Stetattle/Sourdough and Big Beaver populations (Fig. 3). These mountains are Ruby Mountain (2258m) and Pierce Mountain (1516m), respectively. We are unaware of other studies that have identified these mountains as a barrier to dispersal in other organisms, though it is not surprising as both are steep and Ruby Mountain is capped by glaciers.

The Goodell and Newhalem Creek populations that lie to the south and west of Gorge Dam are the most genetically differentiated (Figs. 1, 2). This result could be due to the age of Gorge Dam, which is the oldest of the three dams. However, these results must be interpreted cautiously, as Gorge Dam is only 5 years older than Diablo Dam, and 27 years older than Ross Dam. The earlier establishment of Gorge Dam means that its presence has putatively affected organismal populations for more time as compared to the other dams. Hypothesizing that dam age is correlated with genetic differentiation of populations separated by the dams can be tested by examining F_{ST} (Table 3) and "correct" population assignments (Table 6), where we expect both of these values to be positively correlated with dam age. However, we do not see this pattern in our results because F_{ST} values do not have a linear relationship with dam age.

Another reason for the differentiation of Goodell and Newhalem Creeks from all others is that these two populations are lower in elevation than most of the other sites we sampled (~380m above sea level), with the highest (Panther Creek) at ~830 m. Elevation has been shown to structure amphibian populations in other systems in northwestern North America (Giordano et al. 2007). However, high elevation sites in that study exceeded 1200 m elevation, therefore the difference in elevation across our study site may be too small to detect strong genetic clines across elevation. The population structure observed in *A. truei* could be an analytical artifact; DAPC seeks to maximize the separation between groups while minimizing the variation within groups (Jombart et al. 2010), which means it may find population structure when there is little/weak population structuring. However, our randomization tests indicate that this is not the case, because random population assignments were not able to produce the high assignment probabilities that we obtained for the empirical data (Table 6). Our conclusion is that DAPC was able to detect subtle differences between populations.

These data do not provide a definitive answer regarding the number of populations contained in our datasets. The DAPC provides nearly equal BIC estimates for k values ranging from 1 to 3 (complete dataset; Supplemental Fig. 3), or $k = 3$ to 5 (50% complete dataset; Supplemental Fig. 5), and our Admixture results suggest a single population. The higher k value selected for the 50% complete dataset in DAPC analyses is likely due to the increased number of loci, and therefore increased information content in the dataset (at the expense of increasing the level of missing data per locus), for this dataset. However, how missing data and dataset size affect the ability of the DAPC implemented in Adegenet to detect the true number of

populations needs to be addressed in a simulation study and is outside of the scope of this paper.

The lack of clear geographic structure in our de novo analyses could be due to a few reasons. First, the structure detected in the DAPC analyses could be due to sex-linked loci. Although early cytogenetic research indicates that *Ascaphus* does not appear to have sex chromosomes (Wickbom 1950), recent research has found evidence for sex-linked loci across anuran genomes in autosomal chromosomes (e.g., Brelsford et al. 2013), and the presence of homomorphic sex chromosomes (Brelsford et al. 2016). With limited *Ascaphus* genomic resources available, it is difficult to test if the genetic structuring we see is actually due to sex-linked loci, but this hypothesis cannot be ruled out. Second, because of the high mortality before metamorphosis in this species, the vast majority of individuals that we sampled were larvae ("tadpoles"; 178/183). Given the larval-biased composition of our dataset, our results might have revealed geographic structure if we sampled more adults. Third, lab contamination or bioinformatic errors could have wiped out any geographic structure in the data. Using replicated data collection strategies using separate sequencing lanes can provide important checks on the reliability of RADseq experiments. Finally, the lack of population structure could be a reflection of recent population subdivision. These populations are likely in early stages of divergence, and we do find evidence of population structure associated with the oldest dam in this system.

Missing data levels and dataset size

Missing data had an effect on the outcome of some of our analyses. In pyRAD (Eaton 2014), the user can modulate dataset size for a given number of individuals (in part) by changing the missing data threshold. For instance, allowing a locus (SNP) to be retained in our dataset if it has a minimum of 9/183 individuals present (~5% complete, 95% incomplete) results in a dataset size of 9767 loci, whereas a dataset composed of only 100% complete loci (183/183 individuals at each locus) results in a dataset size of 211 loci (Supplemental Fig. 9). Thus, having more SNPs/loci comes at the cost of increasing the level of missing data. Altering the amount of missing data at a locus is not expected to change results of AMOVA or F_{ST} analyses, but it will increase the variance about these estimates (J. Felsenstein, pers. comm.).

We were able to see the effects of changing the level of missing data in our analyses. With DAPC analyses, $k = 3$ or 4 (BIC difference of 0.15 points; Supplemental Fig. 5) is selected with the 50% missing dataset, in contrast with $k = 2$ for the 100% complete dataset. The level of missing data or other aspects of the data (such as the analysis of sex-linked loci that were discussed above) might play a role with these differences. However, the difference is likely due to the increase in information content when more SNPs are included in the dataset. More studies are needed that examine the effect of missing data on population clustering methods.

Conclusions

In this study, we examined the genetic structure and population connectivity of *Ascaphus truei* in the North Cascades National Park Service Complex using a large SNP dataset. We specifically tested whether hydroelectric dams and their associated lakes are correlated with population structure in this species. We found evidence for high levels of population connectivity throughout the North Cascades National Park System Complex through various analyses. Additionally, we were able to detect subtle population structure at a fine geographic scale that coincides with the oldest dam (Gorge Dam), suggesting that SNP data are able to detect recent population structure that may have elapsed over as few as 12 generations from when the Gorge Dam was erected.

Acknowledgements We graciously thank the Wildlife Research Program from Seattle City Light for funding. Ron Tressler with Seattle City Light and Becky Johnson in the University of Washington Department of Biology helped coordinate logistics and project administration. Special thanks to M. Miller, I. Caviedes-Solis, A. Cortes, A. Gottscho, N. Herschberger, B. Peecook, H. Rockney, P. Tosello, D. Vilhena, and J. Wilkey for assistance with fieldwork. We wish to thank National Park Service personnel A. Rawhouser and B. Kuntz for aid in study design, R. Rochefort with acquisition of collecting permits, and North Cascades National Park superintendent P. Jenkins for supporting this research. Thanks to the Leaché Lab group and two anonymous reviewers for critical and thoughtful comments that greatly improved the manuscript. Lastly, we thank B. Rannala for his assistance with BayesAss and D. Petkova for her help with EEMS analyses. This work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. The completion of this work was aided by a pre-doctoral fellowship from the National Institutes of Health given to JAG.

Author Contributions JAG and ADL conceived and designed the project. JAG completed the fieldwork, lab work, analyses, and wrote the manuscript. ADL provided lab resources, helped with analyses, and edited the manuscript.

Data Accessibility DNA Sequences are available on NCBI's Short Read Archive, BioProject ID PRJNA362634.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest to declare. Furthermore, study design, data collection and analyses, and de-

cision on where to publish this manuscript, were not influenced by the sponsors of this work.

References

Alexander D, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664

Benke AC (1990) A perspective on america's vanishing streams. J N Am Benthol Soc 77–88

Björklund M, Bergek S (2009) On the relationship between population differentiation and sampling effort: is more always better? Oikos 118(8):1127–1129

Brelsford A, Dufresnes C, Perrin N (2016) High-density sex-specific linkage maps of a european tree frog (hyla arborea) identify the sex chromosome without information on offspring sex. Heredity 116(2):177–181

Brelsford A, Stöck M, Betto-Colliard C, Dubey S, Dufresnes C, Jourdan-Pineau H, Rodrigues N, Savary R, Sermier R, Perrin N (2013) Homologous sex chromosomes in three deeply divergent anuran species. Evolution 67(8):2434–2440

Brown HA (1975) Temperature and development of the tailed frog, Ascaphus truei. Comp Biochem Physiol A 50(2):397–405

Bury, R. B. and M. J. Adams, 1999. Variation in age at metamorphosis across a latitudinal gradient for the tailed frog, *Ascaphus truei*. Herpetologica 283–291

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes Genomes Genet 1(3):171–182

Claussen DL (1973) The water relations of the tailed frog, Ascaphus truei, and the pacific treefrog, hyla regilla. Comp Biochem Physiol A 44(1):155–171

Corn PS, Bury RB (1989) Logging in western oregon: responses of headwater habitats and stream amphibians. For Ecol Manag 29(1):39–57

Curtis JM, Taylor EB (2003) The genetic structure of coastal giant salamanders (Dicamptodon tenebrosus) in a managed forest. Biol Conserv 115(1):45–54

Duda JJ, Freilich JE, Schreiner EG (2008) Baseline studies in the elwha river ecosystem prior to dam removal: introduction to the special issue. Northwest Sci 82(sp1):1–12

Dudley N, Baldock D, Nasi R, Stolton S (2005) Measuring biodiversity and sustainable management in forests and agricultural landscapes. Philos Trans R Soc B 360(1454):457–470

Eaton DA (2014) Pyrad: assembly of de novo radseq loci for phylogenetic analyses. Bioinformatics 30(13):1844–1849

Edgar RC (2004) Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. PLOS ONE 6(5):e19379

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011) Local de novo assembly of rad paired-end contigs using short sequencing reads. PLOS ONE 6(4):e18561

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14(8):2611–2620

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. Mol Ecol Resour 10(3):564–567

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annual Rev Ecol Evol Syst 487–515

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol Biol Evol 23(3):691–700

Funk WC, Blouin MS, Corn PS, Maxell BA, Pilliod DS, Amish S, Allendorf FW (2005) Population structure of columbia spotted frogs (Rana luteiventris) is strongly affected by the landscape. Mol Ecol 14(2):483–496

Giordano AR, Ridenhour BJ, Storfer A (2007) The influence of altitude and topography on genetic structure in the long-toed salamander (ambystoma macrodactulym). Mol Ecol 16(8):1625–1637

Hanks EM, Hooten MB (2013) Circuit theory and model-based inference for landscape connectivity. J Am Stat Assoc 108(501):22–33

Harrison HB, Saenz-Agudelo P, Planes S, Jones GP, Berumen ML (2013) Relative accuracy of three common methods of parentage analysis in natural populations. Mol Ecol 22(4):1158–1170

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of snps and microsatellites for parentage and kinship assignment in a wild sockeye salmon (oncorhynchus nerka) population. Mol Ecol Resour 11(s1):150–161

Jombart T (2008) adegenet: a r package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide snp data. Bioinformatics 27(21):3070–3071

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11(1):94

Jones OR, Wang J (2010) Colony: a program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour 10(3):551–555

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci 90(9):4087–4091

Murphy ML, Hall JD (1981) Vaired effects of clear-cut logging on predators and their habitat in small streams of the cascade mountains, oregon. Can J Fish Aquat Sci 38(2):137–145

Nielson M, Lohman K, Sullivan J (2001) Phylogeography of the tailed frog (Ascaphus truei): implications for the biogeography of the pacific northwest. Evolution 55(1):147–160

Nilsson C, Berggren K (2000) Alterations of riparian ecosystems caused by river regulation. BioScience 50(9):783–792

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. Annu Rev Ecol Evol Syst 637–669

Pess GR, McHenry ML, Beechie TJ, Davies J (2008) Biological impacts of the elwha river dams and potential salmonid responses to dam removal. Northwest Sci 82(sp1):72–90

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest radseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species. PLOS ONE 7(5):e37135

Petkova D, Novembre J, Stephens M (2016) Visualizing spatial population structure with estimated effective migration surfaces. Nat Genet 48(1):94–100

Pringle CM, Freeman MC, Freeman BJ (2000) Regional effects of hydrologic alterations on riverine macrobiota in the new world: tropical-temperate comparisons. BioScience 50(9):807–823

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Rambaut A, Drummond A (2007) Tracer v1.5. computer program and documentation distributed by the authors

Richter BD, Mathews R, Harrison DL, Wigington R (2003) Ecologically sustainable water management: managing river flows for ecological integrity. Ecol Appl 13(1):206–224

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584

Sepulveda AJ, Lowe WH (2009) Local and landscape-scale influences on the occurrence and density of Dicamptodon aterrimus, the idaho giant salamander. J Herpetol 43(3):469–484

Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (Ambystoma tigrinum melanostictum). Mol Ecol 14(8):2553–2564

Spear SF, Storfer A (2008) Landscape genetic structure of coastal tailed frogs (Ascaphus truei) in protected vs. managed forests. Mol Ecol 17(21):4642–4656

Wahbe TR, Bunnell FL, Bury RB (2004) Terrestrial movements of juvenile and adult tailed frogs in relation to timber harvest in coastal british columbia. Can J For Res 34(12):2455–2466

Wang J, Santure AW (2009) Parentage and sibship inference from multilocus genotype data under polygamy. Genetics 181(4):1579–1594

Wickbom T (1950) The chromosomes of Ascaphus truei and the evolution of the anuran karyotypes. Hereditas 36(4):406–418

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163(3):1177–1191